Digital Forensics for Dummies

Libby Axen, Michael Doolan, Mason Goida, Tatiana Gomez, and Peyton Ritchie

SIS 220: Investigative Methodology and Forensic Science

Prof. Kelly Crockett

April 26, 2023

Abstract

This paper goes into detail on the history and judicial aspects of digital forensics, along with information about how the product fits into the Daubert compliance and rule 702. The research question was how to automate and simplify digital investigations for low-income and digitally illiterate law enforcement wishing to pursue cybercrime. We hypothesize that if we create an easy to use and simplified digital forensic software, then digital forensics will be accessible to any law enforcement agency and useable by people with different skill sets. We used a variety of open-source tools to create our product. These Linux command-line tools help to image, analyze, and extract information from a device needed for an investigation and make it easier to gain the information. The product was tested on a small scale but there are still many necessary improvements to make the product usable. Some of these potential improvements would be adding a memory analysis function, optimization of the program, using a better coding program, adding a graphical interface, make it able to run on multiple operating systems and have better compatibility, as well as making it more user friendly. This study and product creation is important because it will greatly help the speed and ease the process of collecting digital data and make it more accessible to all law enforcement agencies of varying technological ability. According to our results and findings, the creation of a product like this one is possible and will improve the field of digital forensics.

Introduction

The question is, how to automate and simplify digital investigations for low-income and digitally illiterate law enforcement wishing to pursue cybercrime. The purpose of this software is to do just that by creating free, easy-to-use software that automates and simplifies the investigative process of a computer. Ideally, low-budget and digitally illiterate law enforcement agencies can use our software with minimal knowledge of cyber-related disciplines. This paper will show the research process on the history and judicial aspects of digital forensics as well as explain how this problem can be solved with the software being used. The problem will be handled by creating a product that uses a variety of tools that simplify the digital forensics process. These tools will use techniques such as extracting and imaging data, and steganography.

Literature Review

As part of this product and our group's presentation, we will conduct thorough research into the history and judicial process of digital forensics. This includes identifying past cases, the evolution of the technology, and how our product qualifies with Daubert and Rule 702.

History

Digital Forensics has been around since the 1980's however grew in popularity in 1999 when it was more intricate and useful (Garfinkel, 2010). According to Interpol, digital forensics is the process of extracting data from electronic evidence and processing it into actionable intelligence to present the findings for prosecution. In an infographic from the University of Nevada, Reno, nine phases of digital forensics are provided: responding, seizing the devices, properly collecting the evidence, securing it, acquiring, analyzing, assessing the data, documentation, reporting after the investigation, and expert witness testifying. There is a very specific way to perform digital investigations and collect the evidence without destroying or

tampering with it. Autopsy is the most common application that is used as a digital forensic tool. An example of an investigation utilizing digital forensics was in 1999 when a teenage kid from Florida penetrated computers from the Department of Defense and NASA. Jonathan James was recognized by many with his hack at such a young age, he became the first juvenile hacker to be sent to prison (Gary Cohen, 2021). Digital forensics can be used to solve many different types of crimes such as fraud cases, cybercrimes or any other crimes that require pulling evidence from the internet.

According to "The future of the forensic science providers – Time to re-think our structures?" it states, "The way that the forensic laboratories are operating within the police and the criminal justice system hasn't evolved much during the last decades…" (De Kinder & Pirée, 2020). This shows how an update and new method is needed to make digital forensics more accessible. The article mentions how our society and crime has become very digital and it and forensic science needs to keep up. It is also mentioned how in the future "it is important that there is only a single structure of forensic service providers present in a country, which all report to a single centralized unit" (De Kinder & Pirée, 2020). Which goes to support the need for our product.

Crimes and criminal investigations are merging into the digital era. As technology is becoming more advanced, more information is being stored online that may benefit the accuracy of a case. For example, BTK Killer Dennis Rader tortured and killed his victims in Wichita, Kansas. Rader was also famous for taunting police officers by sending them evidence while he was still at large. His last taunt occurred when he sent a floppy disk to the local Fox-TV News Station which was later examined using digital forensics. This led to Rader's arrest as the disk

contained a forgotten Microsoft Word document, containing personally identifiable information (PII) which the police used to find his doorstep.

Judicial

Some ethical concerns regarding the search and seizure of computer data are the loss of privacy due to less restrictions on the confiscation of digital data, allowing investigators to go beyond the scopes of their investigation and intrude on other nonresponsive data. Because of this, many people believe that digital forensics is not as admissible in court due to the possibility of violating the 4th amendment. Traditionally, there is a two-stage process for collecting digital evidence where "agents enter the physical place to be searched and seize all computers. Second, agents conduct an electronic search for the responsive data described in the warrant" (Texas Tech Law Review, 2015). During the search for data, it is possible for agents to accidentally recover nonresponsive data that is not included in the warrant. If this occurs, it is possible for all the evidence to be unusable as it breaks the confines of the warrant.

Furthermore, there are multiple issues of admissibility of digital evidence in the court room. Due to the nature of gathering digital evidence, "factors such as inadequate chain of custody, not maintaining legal procedures and inadequate evidential integrity" (Yeboah-Ofori & Brown, 2020) have all caused inadmissibility in presenting this specific type of evidence in court. Previously, little to no official guidelines on evidence collection for digital crimes have been established, especially as it is a newer field of forensic investigation. As time has progressed and cybercrimes have grown more relevant, guidelines revolving around the collection of evidence have been developed by noticeable organizations such as INTERPOL, FBI, Office of Justice, and many notable universities, however, there seems to be no universal

standard on the subject of digital evidence collection. Due to this, the admissibility of digital evidence is questionable due to the multiple different standards from differing organizations.

There exists a myriad of laws involving various topics such as data protection, privacy, and hacking inside the United States, both on the federal and state level. The federal Computer Fraud and Abuse Act ("CFAA"), 18 U.S.C. § 1030, is the primary statutory mechanism for prosecuting cybercrime (Global Legal Group, n.d.). While this statue applies to a wide range of cybercrime, CFAA is used to prosecute illegal system breaches and cyber extortion.

Daubert and Rule 702

The Daubert standard is used to assess if a testimony by an expert witness is based on valid and scientific reasoning and can be applied to the facts of the case (Cornell Law, 2023). Rule 702 requires an expert to be qualified, testimony addresses a subject matter where the factfinder can be assisted by an expert, the testimony to be reliable, and the testimony to fit the facts of the case (Cornell Law, 2023). All the tools that are used in our software are open source and well-used in both government and industry. This product does not aim to create new state-of-the-art digital forensics tools, but instead automate already well-respected digital forensics tools. Because our software is simply a compacted and automated collection of other well-tested tools, we believe that this product fits within the requirements of Daubert and Rule 702 and thus is acceptable for use in court, under the assumption that our tool is polished, functional, well documented, and provides no bias or errors in collecting evidence. When using this tool, it may be required that an expert on digital forensics (and eventually an expert on this tool) testifies to validate the tool and its findings in a court of law. This way the evidence that is found can be used and strong enough to support the findings of the case.

Methodology

Digital forensics is becoming a vital part in law enforcement agencies, especially with the emergence of new technologies. This field is not only beneficial for law enforcement, but also for suspects and other attackers. A trained official or digital forensics professional has become a desired role in most agencies. With the expertise to combat and investigate digital crimes, making their role of great value. Although, with digitally illiterate officials it has been a struggle for some agencies to adapt to new technologies. This is where our group thought of simplifying the process to assist the law enforcement agencies that have not adapted to new technologies. We hypothesize that if we create an easy to use and simplified digital forensic software, then digital forensics will be accessible to any law enforcement agency and useable by people with different skill sets.

Product Summary

This program is a Linux Shell script that takes a forensically sound image of a computer and automates a variety of Linux command line tools on that image. The program starts with a verification check that the user is running the program with root privileges.

Second, the program installs the required tools and dependencies necessary to run the program. Specifically, the program installs Pip, Git, Python 3.10 and its dependencies, Python 2.7 and its dependencies, Java and its dependencies, Volatility, Autopsy, Sleuthkit, Foremost, bulk-extractor, Hashdeep, browser-history, and Steghide. Pip and Git are package management tools, Python and Java are programming languages, Volatility is a memory forensics and analysis tool, Autopsy, Sleuthkit, bulk-extractor, Foremost, and browser-history are disk image analysis tools, Steghide is a reverse steganography tool, and Hashdeep is a hashing/integrity tool.

Next, the program requires the user to create and name a case file to store the results of the program. Following the tool installation and case path functions, the program requires the

user to connect a storage device to the host computer. The program will then take a forensically sound image of the storage device using the inbuilt dd command in Linux and store it on whatever storage device the user selects. This image will be the basis for all of the disk image analysis tools and the reverse steganography tool. The program will then hash the image file, which is used in digital forensics and even cybersecurity at large for data integrity and verification. Attempting to change or write anything on the image will result in a different hash. This program will check the hash periodically and will alert the user if the hash function has changed.

This program will then continue with bulk-extractor, a disk image analysis tool that scans the target image without parsing the image. Specifically, the tool extracts images, pdfs, pngs, credit card numbers, phone numbers, email addresses, social security numbers, social media accounts, search history. GPS and locations store the output to a summary file called summaryfile.txt, where the user can then perform other searching mechanisms to look for specific information. Additionally, all images, pngs, and pdfs are sent to a separate directory called “bulk_extractor_images” where they can be processed by Steghide.

Following bulk-extractor, the program runs a tool called Foremost, a Linux tool that scans an image and recovers any deleted files that have not been overwritten by the operating system. The results of the tool are stored in a directory called “foremost_results”, where the user can browse any deleted files including pictures, documents, and other files necessary to an investigation. The images, pngs, and pdfs are sent to the bulk_extractor_images directory to be further processed with Steghide.

The program will continue with a tool called Steghide, a reverse steganography tool that scans all images, pngs, and pdfs for any evidence of embedded data and processes the pictures in

its own, unique directory. It uses the bulk_extractor_images directory as its input, and recursively traverses the directory for any images that contain hidden data. It is important to note that Steghide does not work on pngs due to their unique nature, and a separate tool must be used on pngs to be able to perform reverse steganography on them.

To conclude, the program runs a tool called browser-history that scans an image and pulls all the browser history off the computer and displays it in a .csv document, which is an excel spreadsheet extension. With this, the browser history can be processed in a separate tool such as Excel for further analysis.

Our group chose these tools because we believe they would be the most beneficial for simplifying and automating a standard digital forensics investigation. The platform consists of tools that are vital to a digital forensic investigation such as file analysis, timeline analysis, hashing, deleted file recovery, and more.

Testing

After developing and bug testing the program, our group ran the digital forensics for dummies program on two storage devices, one a 128-gigabyte flash drive and one 4-terabyte spinning hard drive, both allegedly brand new from Amazon and Costco respectively. Our group chose a small flash drive and a large storage device to ensure that our product worked on a variety of storage devices and that size or brand would not break our program. To simulate a standard digital forensics investigation, our group used steganography to place pictures of various babies and toddlers inside a picture of a hot air balloon. Additionally, our group also created and deleted several random text files on each storage device. This is to simulate the use of steganography to hide child pornography and to simulate what happens when a suspect deletes a file on a Windows machine. Our group also used precompiled examples for demonstration

purposes, as the program takes an extremely long time to run, approaching almost 13 hours for the 4-terabyte storage device.

Results

The program was able to successfully image and analyze both storage devices successfully, but the results were disturbing. It turns out that both storage devices were not brand new as advertised. Our group learned that the brand-new storage devices were actually recycled, used storage devices. We found this out because our program was able to successfully analyze and recover not only the deleted text files and hidden pictures that we placed but also all the previous owner's files and display them in the results. It turns out that the previous owners stored sexually explicit images of various males and females on their storage devices, and as a result, our group elected to delete the entire directory and wipe the storage devices clean approximately 4 times. Unfortunately, because of us deleting the results, we only showed snippets of the results to the recruiters at the forensics fair to avoid being inappropriate. In total, the program took approximately 2 hours to run the 128GB flash drive and approximately 13 hours to run the 4TB HDD. This is due to the invasiveness and the thoroughness of the tools we used, along with the sheer space of the image files.

Discussion

After developing and testing our product, our group believes that the product was successful with our goal of simplifying the digital forensics process for low income and digitally illiterate law enforcement agencies. The program is free and open source, soon to be posted on Github as a public repository. Additionally, the product was successfully tested and demonstrated at the forensics fair, with plenty of positive feedback from attendees. The product was able to successfully image and analyze a storage device and the product ensured that the

appropriate evidence was collected in a forensically sound manner. With this, our group was successful in automating the digital forensics process and simplifying the workload for a typical law enforcement official via a Linux command line tool.

There were some complications, however, such as our tool analyzing a device that had a previous owner. This was not our intention, as recovering files that are expected to be removed and deleted is an invasion of privacy. Even more so, our results showed sexually explicit images from the previous owner after using a file recovery tool called Foremost on the flash drive and the hard drive, which made us delete the results of our findings to avoid being inappropriate. While this would have demonstrated the results and effectiveness of our tools, it is inappropriate and wrong to keep sexually explicit images of people without their consent. As a result, our group elected to delete and wipe the storage devices, thereby destroying our results.

Despite our success, the program is far from perfect. The program lacks critical functions in a standard digital forensics' investigation, such as memory analysis and presentation functions. Our tool is more accurately a disk image analysis tool than a digital forensics tool. Some of the main features our group intends to add or modify in the future includes memory analysis, optimization, rewriting in a better coding language, adding a graphical interface (GUI), making it able to run on multiple operating systems (Windows, Linux, macOS), have better compatibility, and making it more user-friendly.

Though the product was successful in testing it could have been optimized to better fit the time of a law enforcement official and their case. For example, in the initial testing, the product took almost 15 hours to run to successfully image and analyze the storage devices. The standard law enforcement agency does not have this much time to analyze this data without risking the exposure of evidence. Another example includes how memory analysis has become a large part

of the forensics world. This is where instead of performing analysis on storage such as USBs and hard drives, analysis will be conducted on the RAM or the memory of a computer at a specific point in time. The analysis could provide additional insight and introduce more tools to automate the digital forensics process.

A visible pattern can be seen upon examining the improvements that our group intends to make to our product. If we can combine all our opinions into one choice, we would decide to improve the user experience of our product. The original idea for our project was to create a "point and click" design that will be as simple as clicking a button to automate a task in forensics. Although this is still our idea, it was quickly realized that complete automation and simplification is simply not possible with a field as complex as digital forensics. Overall, an improved user experience and interface would help us reach our goal of simplifying the digital forensics process and assisting those who have no technical skills.

Conclusion

Our findings support our hypothesis that if we create an easy to use and simplified digital forensic software, then digital forensics will be accessible to any law enforcement agency and useable by people with different skill sets. because we were able to make a program that fits the specifications of what we were trying to achieve. The program is a compilation of many different programs that are already ethical and used in trials, therefore it is a cohesive program that will be useful within the field and there is no other program that already exists. There are some more testing and improvements that would be ideal to make before the product gets used in investigations but the foundations in place are key. This research is vital within the field of forensic science and digital forensics as it will help speed up the process, make it more accessible, keep everything ethical and prevent tampering with or losing the evidence.

References

Cohen, G. (2022, August 15). *Throwback attack: A Florida teen hacks the Department of Defense and NASA*. Industrial Cybersecurity Pulse. Retrieved February 24, 2023, from https://www.industrialcybersecuritypulse.com/facilities/throwback-attack-a-florida-teen-hacks-the-department-of-defense-and-nasa/

Cornell Law. (n.d.). *Daubert Rule*. Legal Information Institute. Retrieved March 24, 2023, from https://www.law.cornell.edu/wex/daubert_standard

Cornell Law. (n.d.). *Rule 702. testimony by expert witnesses*. Legal Information Institute. Retrieved March 24, 2023, from https://www.law.cornell.edu/rules/fre/rule_702

De Kinder, J., & Pirée, H. (2020). The future of the forensic science providers – Time to re-think our structures? *Forensic Science International (Online), 316*https://doi.org/10.1016/j.forsciint.2020.110471*Digital Evidence and Forensics.*

Digital Forensics. INTERPOL. (n.d.). Retrieved February 24, 2023, from https://www.interpol.int/en/How-we-work/Innovation/Digital-forensics

Garfinkel, Simson L. "Digital Forensics Research: The Next 10 Years." *Digital Investigation*, vol. 7, 2010, pp. S64 - S73, https://www.sciencedirect.com/science/article/pii/S1742287610000368, doi:10.1016/j.diin.2010.05.009.

Global Legal Group. (n.d.). *Cybersecurity Laws and Regulations Report 2023 USA*. International Comparative Legal Guides International Business Reports. https://iclg.com/practice-areas/cybersecurity-laws-and-regulations/usa

INTERPOL. (2021). *GUIDELINES FOR DIGITAL FORENSICS FIRST RESPONDERS* [PDF]. https://www.interpol.int/content/download/16243/file/Guidelines%20to%20Digital%20Forensics%20First%20Responders_V7.pdf

Texas Tech Law Review. (2015). *Executing Warrants for Digital Evidence: The Case for Use Restrictions on Nonresponsive Data*. Vol. 48, 2015-2016. Retrieved from https://heinonline.org/HOL/P?h=hein.journals/text48&i=7

The phases of Digital Forensics. University of Nevada, Reno. (2022, July 14). Retrieved February 24, 2023, from https://onlinedegrees.unr.edu/blog/digital-forensics/

Yeboah-Ofori, A., & Brown, A. D. (2020). Digital Forensics Investigation Jurisprudence: Issues Of Admissibility Of Digital Evidence. *HSOA Journal of Forensic, Legal & Investigative Sciences*, *6*(1), 1–8. https://doi.org/10.24966/flis-733x/100045